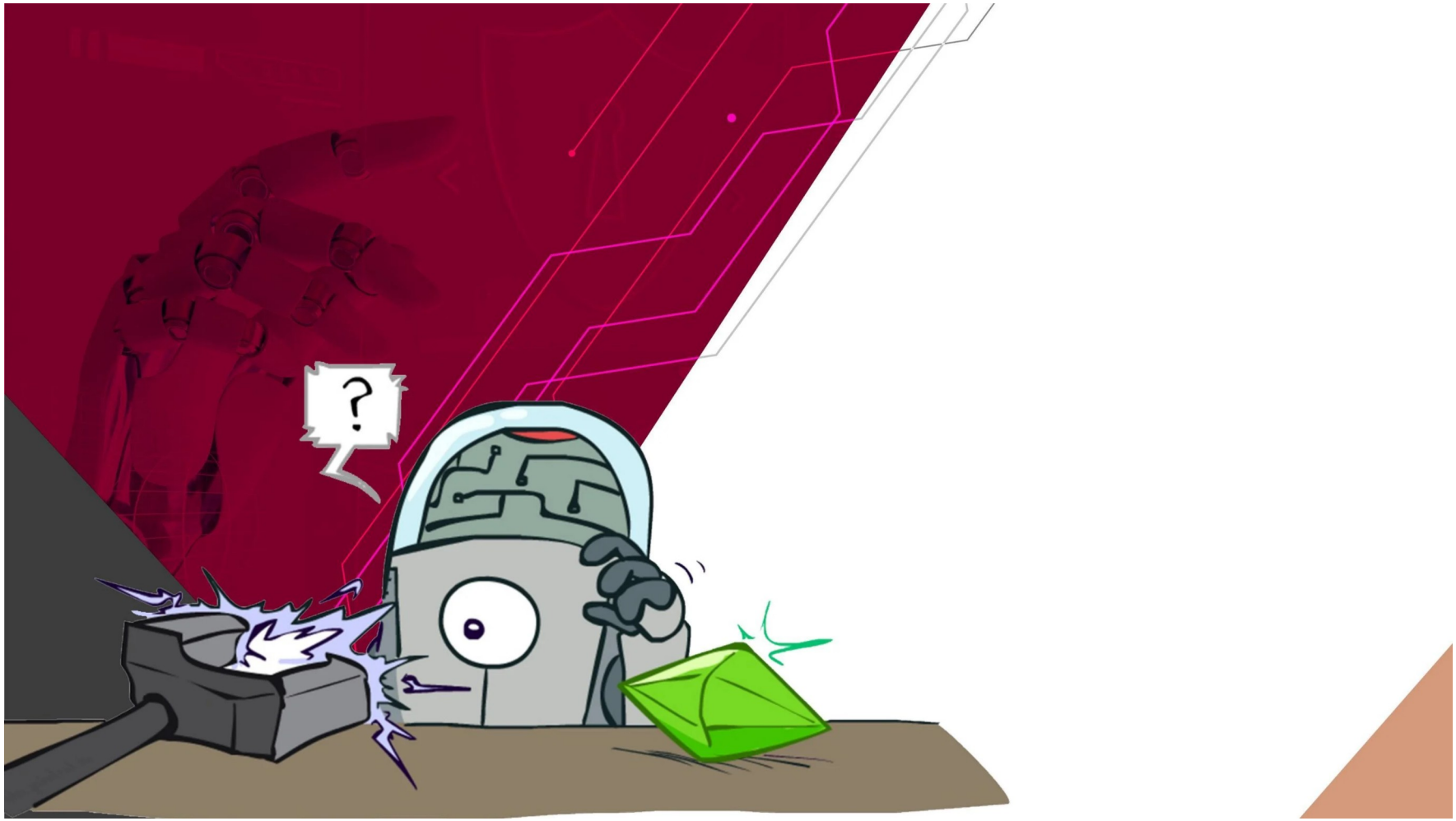
A8. Aprendizaje por Refuerzo

XUNTA DE GALICIA
CONSELLERÍA DE CULTURA, EDUCACIÓN E UNIVERSIDADE

IES de Teis
Avda. de Galicia, 101
36216 – Vigo
886 12 04 64
ies.teis@edu.xunta.es

GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL
Unión Europea-
NextGenerationEU

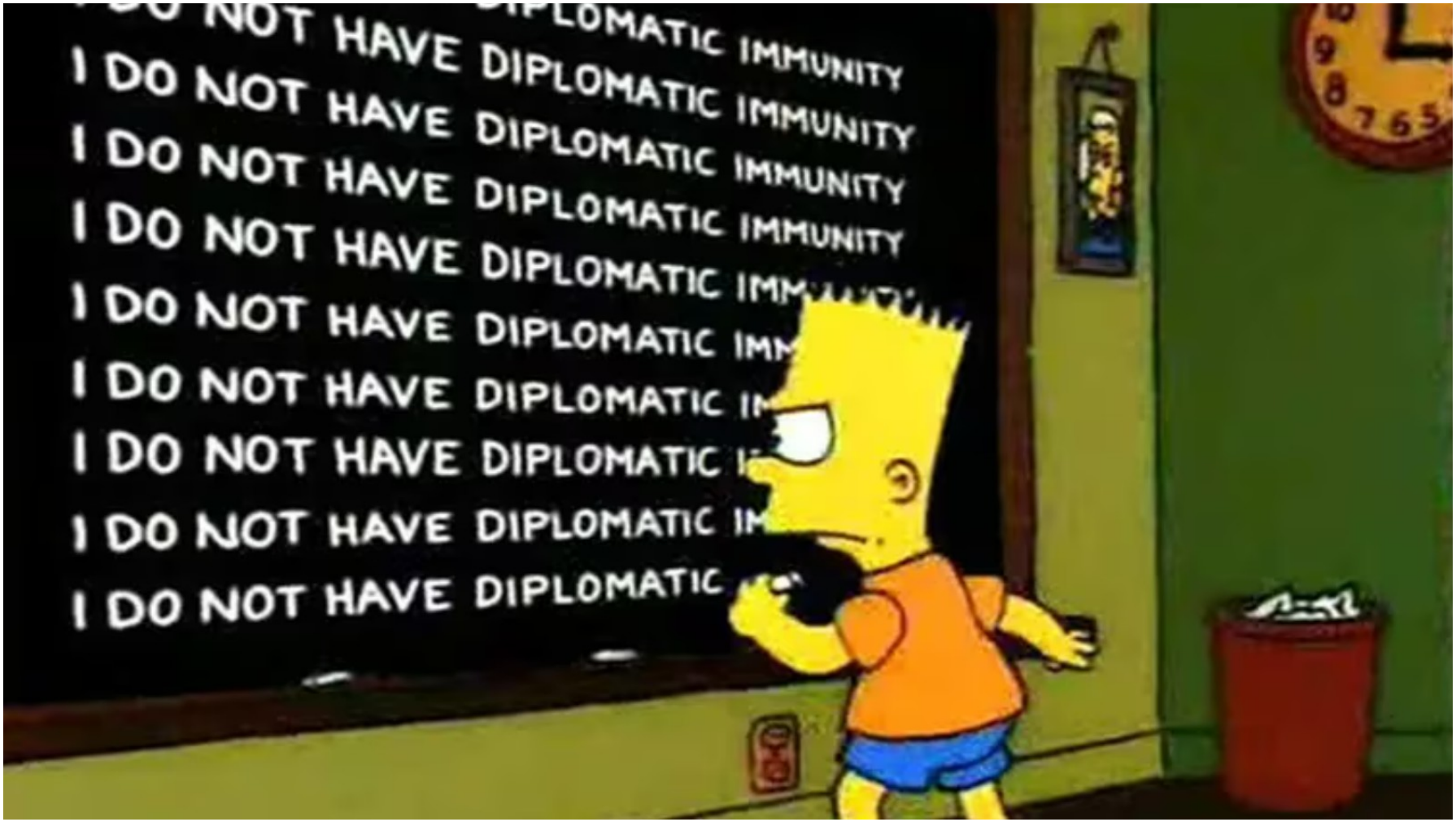


A8. Aprendizaje por Refuerzo

Índice.

A8. Aprendizaje por Refuerzo

1. ¿Qué es el Aprendizaje por Refuerzo?



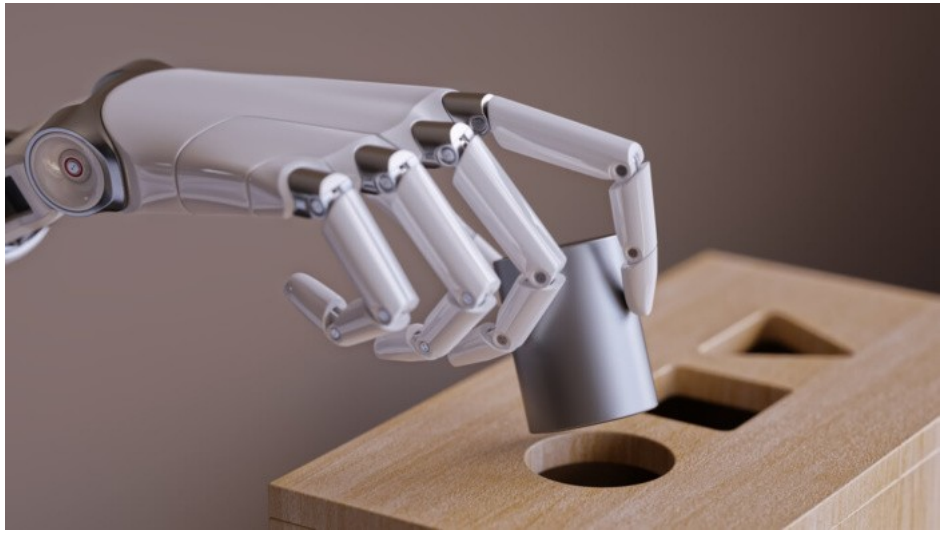
A8. Aprendizaje por Refuerzo

1. ¿Qué es el Aprendizaje por Refuerzo?

El **Aprendizaje por Refuerzo**, en el ámbito del Machine Learning y la Inteligencia Artificial, es un tipo de programación dinámica consistente en el entrenamiento de algoritmos a través de un sistema de recompensas y castigos, buscando maximizar una recompensa acumulativa.

Este tipo de aprendizaje permite a las redes neuronales aprender cómo lograr objetivos complejos o maximizar una dimensión específica a lo largo de múltiples pasos.

El Aprendizaje por Refuerzo no tiene etiquetas de salida (no es de tipo Supervisado) y aunque los algoritmos aprenden por sí mismos (tampoco son de tipo No Supervisado), basan su funcionamiento en un esquema de premios y castigos, en los que cada acción está afectada por múltiples variables que cambian con el tiempo.



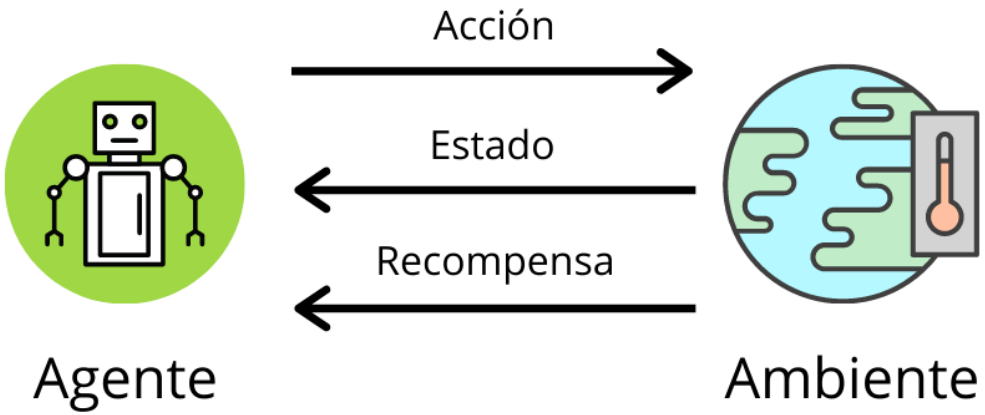
2. Componentes importantes.

El Aprendizaje por Refuerzo propone un nuevo enfoque con el fin de hacer que la máquina aprenda, para lo cual, postula los dos siguientes componentes:

- **Agente:** modelo que se va a entrenar y aprenderá a tomar decisiones.
- **Ambiente:** entorno en el que interactuará el Agente y que tendrá limitaciones y una serie de reglas posibles en cada momento.

Entre estos componentes, hay una relación que se retroalimenta y cuenta con los siguientes nexos:

- **Acción:** conjunto de las posibles acciones que puede tomar en cualquier momento el Agente.
- **Estado (del ambiente):** indicadores del ambiente de cómo están los diversos elementos que lo componen en ese momento concreto.
- **Recompensas (o castigos):** en función de cada acción tomada por el Agente, se puede obtener un premio o una penalización que servirá para orientar al Agente en si lo está haciendo bien o mal.



3. Funcionamiento del Aprendizaje por Refuerzo.

El Agente debe tomar decisiones para interactuar con el ambiente en cada estado en el que se halle.

3. Funcionamiento del Aprendizaje por Refuerzo.

El Agente debe tomar decisiones para interactuar con el ambiente en cada estado en el que se halle.

- Al principio de todo el Agente se halla en 'blanco', es decir, no sabe nada de lo que tiene que hacer ni de cómo comportarse.
- Toma una de las posibles acciones de forma aleatoria e irá recibiendo pistas sobre si lo está haciendo bien o mal, en función de las recompensas: si el Agente elige la opción A y recibe 100 puntos, volverá a elegir la opción A y puede que se estanque en esa única opción.
- Hay que lograr un equilibrio entre explorar lo desconocido y explorar los recursos del ambiente → dilema de exploración/explotación.
- El Agente explorará el ambiente e irá aprendiendo cómo moverse y cómo ganar recompensas (y evitar penalizaciones). Al final, almacenará el conocimiento en una normas llamadas políticas.

4. Características.

Las características más importantes del Aprendizaje por Refuerzo son las siguientes:

- **Ausencia de supervisor:** no hay un supervisor directo que indique las acciones correctas. En su lugar, se utiliza una señal de recompensa o número real para guiar el aprendizaje.
- **Toma de decisiones secuencial:** el aprendizaje por refuerzo implica tomar decisiones en secuencia, donde las acciones realizadas tienen impacto en los datos y resultados futuros.
- **Importancia del tiempo:** el factor tiempo es crucial en los problemas de refuerzo. Las decisiones tomadas en momentos específicos pueden influir en las recompensas obtenidas y en el rendimiento general del agente.
- **Retraso en la retroalimentación:** a diferencia de otros enfoques de aprendizaje, la retroalimentación en el aprendizaje por refuerzo no es instantánea. Existe un retraso entre la acción y la retroalimentación recibida, lo que hace que la toma de decisiones sea más desafiante.

5. Ventajas.

Las ventajas que proporciona el Aprendizaje por Refuerzo son las siguientes:

- Ayuda para determinar en qué situaciones se necesita una acción.
- Permite descubrir qué acciones generan una mayor recompensa a largo plazo.
- Proporciona al Agente de Aprendizaje una función de recompensa.
- Facilita el descubrimiento de los métodos óptimos para alcanzar grandes recompensas.

6. Inconvenientes.

El modelo de Aprendizaje por Refuerzo no se puede aplicar en todas la situaciones. Algunas de las condiciones son las siguientes

- Si no se dispone de suficientes datos para resolver un problema mediante un método de Aprendizaje Supervisado.
- Si no se dispone de mucho tiempo y recursos informáticos.

7. Desafíos.

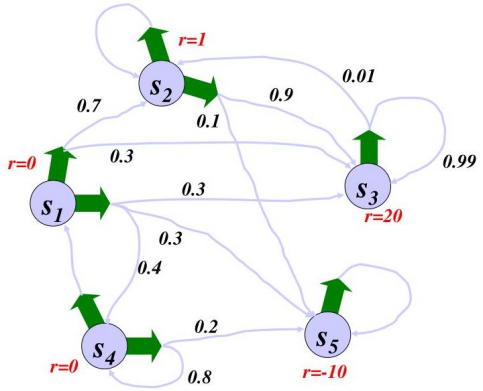
La aplicación del Aprendizaje por Refuerzo se enfrenta a estos desafíos importantes:

- **Diseño de las características/recompensas**: es necesario estar involucrado en el diseño de las características y de las recompensas del sistema, debido a que suelen influir significativamente en los resultados del aprendizaje.
- **Influencia de los parámetros**: los parámetros utilizados pueden afectar a la velocidad de aprendizaje y a los resultados obtenidos, por eso se requiere un ajuste adecuado para obtener un buen rendimiento.
- **Observabilidad parcial:** en entornos realistas, puede haber una observabilidad parcial, es decir, el Agente sólo puede observar una parte limitada del entorno, lo que puede dificultar un toma de decisiones precisa.
- **Sobrecarga de estados:** un exceso de refuerzo puede llevar a una sobrecarga de estados, implicando un gran número de posibles estados del sistema, disminuyendo los resultados del aprendizaje y dificultando la generalización efectiva del Agente.
- **Entornos estacionarios:** los entornos realistas pueden cambiar con el tiempo, conocido como no *estacionariedad*, esto añade un grado de complejidad al aprendizaje, ya que el Agente debe adaptarse a los cambios en el entorno para seguir tomando decisiones óptimas.

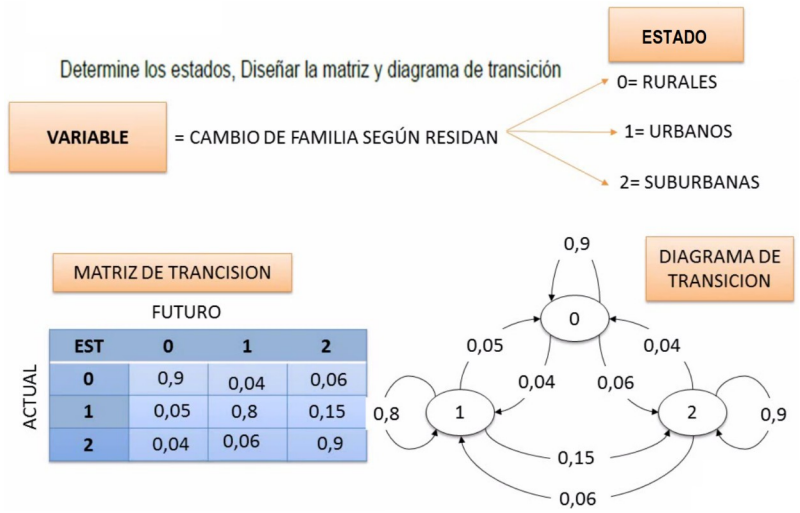
8. Enfoques de la implementación.

Hay dos modelos importantes en el Aprendizaje por Refuerzo:

- **Proceso de decisión de Markov:** método de aprendizaje basado en un enfoque matemático para mapear una solución de aprendizaje.



- **Q-Aprendizaje**: método de aprendizaje basado en valores para proporcionar información en función de la cual el Agente tomará una acción.



9. Tipos de métodos de Aprendizaje por Refuerzo.

Hay dos tipos de métodos de Aprendizaje por Refuerzo:

- **Refuerzo positivo:** si un evento recompensador se asocia con un comportamiento específico. Aumenta la probabilidad y la frecuencia de la conducta y tiene un impacto positivo en las acciones del agente. Ayuda a maximizar el rendimiento y mantener cambios a largo plazo, aunque un exceso de refuerzo puede llevar a la optimización extrema y afectar los resultados.

- **Refuerzo negativo:** si una conducta se fortalece debido a la eliminación o evitación de una condición negativa. Define el rendimiento mínimo requerido y ayuda a evitar consecuencias adversas. Sin embargo, su limitación es que solo proporciona suficiente incentivo para cumplir con el nivel mínimo de rendimiento.

10. Aplicaciones actuales del Aprendizaje por Refuerzo.

El Aprendizaje por Refuerzo tiene numerosas aplicaciones en la actualidad, entre las que cabe destacar:

- Robótica en la automatización industrial.
- Planificación de estrategias empresariales.
- Aprendizaje automático y el procesamiento de datos.
- Creación de sistemas de capacitación que proporcionan instrucción y materiales personalizados según los requisitos de los estudiantes.
- Control de aeronaves y control de movimiento de robots.